Fusion Method For Physical Systems Based On Physical Laws

Nageswara S. V. Rao, David B. Reister, Jacob Barhen
Center for Engineering Science Advanced Research
Computer Science and Mathematics Division
Oak Ridge National Laboratory
Oak Ridge, TN 37831-6355
{raons,reisterd,barhenj}@ornl.gov
Proceedings of Int. Conf. on Information Fusion, 2000.

Abstract - *We consider a physical system described by a set of parameters. Each parameter is either measured by a number of sensors or estimated by a set of computer programs that use sensor measurements. As a result, the resultant parameter values could be widely varying. We propose a fusion method that combines the measurements and estimators based on the physical laws that relate the parameters. In comparison with the traditional fusion problems, there is no training set that provides the actual parameter values. Furthermore, since every parameter is measured or estimated, there are no parameters whose actual values are known. We propose a fuser based on the least violation of the physical laws that relate the parameters. Under certain smoothness conditions on the physical law, we show the asymptotic convergence of our method, and also derive distribution-free performance bounds based on finite samples. We illustrate the effectiveness of this method for a practical problem of fusing well-log data in methane hydrate exploration. For this problem, data fusion method resulted in an order of magnitude improvement in the accuracy compared to the best set of estimators for the key parameter of porosity.*

Keywords: Sensor fusion, physical laws, methane hydrates.

1 Introduction

We consider a physical system wherein each parameter is either measured using an instrument or estimated using a computer program based on the measurements. In general, it may not be possible to know the actual parameter values, since all measurements and estimators based on measurements introduce errors of different types. In particular, there could be both systematic and random errors in the measurements as well as in the estimators. In certain cases, very accurate sensor noise models are available, but such models are not easily obtained for estimators based on complicated computer codes. We address the problem of fusing various measurements and estimators so that for each parameter the fused estimate is superior to any one of the estimators or measurements. There is no real way of knowing the accuracy of various measurements or estimators since the actual values are not known. One of the challenges of this problem is that the traditional pattern recognition or fusion solutions are not applicable since there is no training data. In other problems, the training data can be used to design powerful fusers which can be shown to be at least as good as the best measurement or estimator [12, 13, 11]. Lack of training data makes the available methods inapplicable to the present problem. Indeed, newer problem formulations and solutions are needed to address this issue as shown by us in [16].

There are physical laws that relate the parameters of the physical systems. For example, in a simple mass sliding on a friction-less surface, we have $f = ma$, where f is force, m is mass and a is acceleration. Typically such physical laws are derived from the first principles, and are independently tested. Such laws are possible in non-physical systems too, for example information theoretic laws in mathematical systems or computer programs, and our method can be applied to such cases. However, physical laws are the most direct ones to use in the proposed method. In general, we consider a set of parameters that are related by a physical law. For each parameter, let us choose a measurement or an estimator. The accuracy of this set of measurements and estimators is determined by how well the physical law is satisfied. In this sense, the "violation" of physical law is an indication of error, and the set of estimators that achieve the least violation of the physical law is the most preferred one to choose.

We show in this paper that the physical laws can be used to design a fuser for each of the parameters. If the sensor error distributions are known, we show that the isolation fusers [11] can be designed to ensure that the performance at least as good as the best set of estimators. In the practical case, where we only have sensor measurements, we show that the above results hold asymptotically (as sample size approaches infinity), under certain Lipschitz properties of the physical law and fusion functions. For finite sample sizes, we show a distribution-free result that given large enough sample size the fuser performs better than the best set of estimators within a specified precision and with a

specified probability in the spirit of Probably Approximately Correct (PAC) learning [17, 19].

In section 2, we define the fusion problem for a physical system, and show how physical laws can be used to design a fuser in Section 3. In Section 4, we discuss the practical problem of fusing the well log data collected in the exploration of methane hydrates.

2 Physical Systems and Laws

Consider that a physical system is specified by the parameters $P(z) = (p_1(z), p_2(z), \ldots, p_n(z))$ with $p_i(z) \in \Re$, where z is one-dimensional variable such as time or position. Each parameter $p_i(z)$ is deterministic in that repeated observations for a fixed z by an error-free (ideal) sensor yield the same value.

Each parameter p_i is measured by a_i instruments and estimated by b_i estimators such that $a_i \geq 0$, $b_i \geq 0$, and $a_i + b_i \geq 1$. The measurements corresponding to $p_i(z)$ are denoted by

$$m_i(z) = \{m_{i,1}(z), m_{i,2}(z), \ldots, m_{i,a_i}(z)\}$$

and the corresponding estimators are denoted by

$$e_i(z) = \{e_{i,1}(z), e_{i,2}(z), \ldots, e_{i,b_i}(z)\}.$$

The measurements are obtained by instruments such as sensors, and estimators are obtained using computational procedures that depend on the measurements of possibly other variables. Thus, there are $a_i + b_i$ competing values for each parameter, and in general we do not know which one is more accurate. The measurements are assumed to be noisy in that repeated measurements by a sensor of $p_i(z) = x$ for a fixed value are distributed independently according to the distribution $P_{m_{i,j}|x}$, which is denoted by $P_{m_{i,j}|p_i(z)}$. Again, we consider that the measurement error distributions are not dependent on z, and they depend on the actual value x of the parameter. Thus, $m_{i,j}$ is a random variable. The estimator $e_{i,j}$ is a (deterministic) function of the measurements, and hence is also a random variable. The joint distribution of the measurements is denoted by $P_{m_1,m_2,\ldots,m_n|p_1,p_2,\ldots,p_n}$.

There is a physical law

$$L[p_1(z), p_2(z), \ldots p_n(z)] = 0$$

which relates the actual parameters corresponding to z. We assume that $L[.]$ satisfies the reasonable *monotonicity* condition: for any y_1, y_2, $|y_1| \leq |y_2|$, we have

$$\begin{aligned}&|L[p_1(z), \ldots, p_i(z) + y_1, \ldots, p_n(z)]| \\ &\quad \leq |L[p_1(z), \ldots, p_i(z) + y_2, \ldots, p_n(z)]|.\end{aligned}$$

Monotonicity means that accurate parameter estimators yield no lesser "magnitude" of violation of the law compared to less accurate estimators. Note that $L[.]$ itself does not dependent on z, i.e. informally speaking, we only consider time-invariant or position-invariant laws in that the law itself is not dependent upon z. Such laws are not guaranteed to exist, especially for complicated dynamical systems, but good approximations exist for smaller time scales. Also, if multiple laws of the form $L[.] = 0$ exist that only relate certain subsets of parameters, they can be added to obtain one that relates all p_is.

If we choose a single estimator or measurement $\hat{p}_i$ for the parameter, the closeness of $L[\hat{p}_1(z), \hat{p}_2(z), \ldots, \hat{p}_n(z)]$ to 0 determines how closely the law is satisfied. Let a *basic set*, denoted by S, be a set of measurements and estimators such that for each parameter we choose precisely one measurement or estimator (but not both). The total error due to S is given by

$$\hat{E}(S) = \sum_z L[\hat{p}_1(z), \hat{p}_2(z), \ldots, \hat{p}_n(z)],$$

where $\hat{p}_i$ is the estimator or measurement for p_i. In all there are $\prod_{i=1}^{n}(a_i + b_i)$ possible basic sets, and $\hat{S}$ be the one with least error such that

$$\hat{E}(\hat{S}) = \min_S \hat{E}(S).$$

The expected error of S is denoted by

$$E(S) = \sum_z \int L\left[\hat{p}_1(z), \hat{p}_2(z), \ldots, \hat{p}_n(z)\right] dP_{m_1,\ldots,m_n|p_1,\ldots,p_n},$$

and let S^* be the one with the least expected error such that

$$E(S^*) = \min_S E(S^*).$$

Note that S^* minimizes the expected error but $\hat{S}$ in general does not. If the sensor distributions are not known, then S^* can not be computed, and often $\hat{S}$ is used as a good approximation as in the case of empirical risk minimization [19].

Example 1: Consider a simple illustrative example of known mass m subjected to a constant force f in a friction-free environment. The physical law in this case is $f = ma$, and hence we employ $L[f, a] = (f - ma)^2 = 0$, where a is acceleration and f is force. Let $p_1(z) = f$ and $p_2(z) = a$. Consider that we are given a sensor that measures force and two sensors that measure acceleration. The force measurements are given by $m_{1,1}(z) = f + \epsilon$, for some deterministic ϵ, i.e. the force sensor has simple bias error. The acceleration measurements are given by $m_{2,1} = a + \delta$, $m_{2,2} = 0.7a$, where δ is a small normally distributed error. Then, we have

$$L[m_{1,1}, m_{2,1}] = (\epsilon - m\delta)^2$$

$$L[m_{1,1}, m_{2,2}] = (\epsilon + 0.3ma)^2$$

Consider that $a > 0$, and $\epsilon \geq m\delta$. Under the condition $|\delta| \leq |0.3a|$, we have

$$L[m_{1,1}, m_{2,1}] \leq L[m_{1,1}, m_{2,2}],$$

i.e. for large values of a, the better choice is $m_{2,1}$, otherwise $m_{2,2}$ is a better choice. □

There is a significant leeway and variability in the expression for the physical law, which determines the exact value of the error. In particular, the error can be arbitrarily scaled (up or down) by multiplying $L[.]$ with a suitable constant. Domain-specific knowledge must be utilized in choosing the law as well as the form in which the law is expressed. In addition, there are inherent limitations to the utilization of physical laws in error computation from a theoretical viewpoint. In extreme cases, the errors of various parameters can cancel each other so that

$$L[\hat{p}_1(z), \hat{p}_2(z), \ldots, \hat{p}_n(z)] = 0.$$

In the above example, if $m_{1,1} = f + \epsilon$, and $m_{2,1} = a + \epsilon/m$, then $L[m_{1,1}, m_{2,1}] = 0$. While this is a possibility depending on the form of $L[.]$, such case needs a very cooperative error process for the estimators. Since the individual measurements and estimators are based on different methods such case is very unlikely, especially for systems with a large number of parameters and a complicated form for $L[.]$.

3 Data Fusion Based on Physical Law Violations

A *fusion function* $f_i \in \mathcal{F}_i$ for parameter p_i combines the measurements and estimators such that $f_i(m_i(z), e_i(z))$ is an estimate of $p_i(z)$. Let $f = (f_1, \ldots, f_n)$ denote the *fuser* for all parameters. The expected error due to the fused estimate is

$$E(f) = \sum_z \int L\left[f_1(m_1(z), e_1(z)), \ldots, f_n(m_n(z), e_n(z))\right] dP_{m_1,\ldots,m_n|p_1,\ldots,p_n},$$

and let $f^* \in \mathcal{F}_1 \times \ldots \times \mathcal{F}_n$ be the one with the least expected error. In general $E(f)$ cannot be computed if the error distributions are not known, and hence f^* is not computable. In stead, we compute $\hat{f}$ that minimizes the empirical cost given by

$$\hat{E}(f) = \sum_z L\left[f_1(m_1(z), e_1(z)), \ldots, f_n(m_n(z), e_n(z))\right],$$

based on a set of measurements (also called the sample)

$$\{<(m_1(z), e_1(z)), \ldots, (m_n(z), e_n(z))>: z = 1, \ldots, s\}.$$

Now we discuss methods that ensure $E(f^*) \leq E(S^*)$, and more importantly based on a computable $\hat{f}$ that

$$E(\hat{f}) < E(\hat{S}),$$

with a specified probability based entirely on the measurements and without any knowledge of the underlying distributions.

We now present some preliminaries needed for the main result. For any function $g : [-A, A]^d \mapsto \Re$, let

$$\| g(r) \|_\infty = \sup_{r \in [-A,A]^d} |g(r)|.$$

A function $g(y) : [-A, A]^d \mapsto \Re^a$ is *Lipschitz* with constant k_g if for all $y_1, y_2 \in [-A, A]^d$, we have

$$\| g(y_1) - g(y_2) \|_\infty \leq k_g \| y_1 - y_2 \|_\infty .$$

For example, the sigmoid neural networks are Lipschitz with the constant specified by the parameters of the network [14].

The *covering number* $N_\infty(\epsilon, \mathcal{G})$ of a function class $\mathcal{G}$ is the smallest cardinality for a subclass $\mathcal{G}^* = \{g^*\}$ of $\mathcal{G}$ such that

$$\min_{g^* \in \mathcal{G}^*} \| g - g^* \|_\infty \leq \epsilon$$

for each $g \in \mathcal{G}$. Lipschitz functions defined on bounded domains have easily computable cover sizes. By using a mesh with points ϵ apart, we can cover $[-A, A]^d$ with $\left[\frac{2A}{\epsilon}\right]^d$ points. Then, because of the Lipschitz property of functions of $\mathcal{G}$, we have

$$N_\infty(\epsilon, \mathcal{G}) \leq N_\infty\left(\frac{\epsilon}{k_g}, [-A, A]^d\right) \leq \left[\frac{2Ak_g}{\epsilon}\right]^d.$$

We now discuss the *isolation property* [15] of the fuser class that ensures that the fuser is at least as good as the best basic set. A fuser class

$$\mathcal{F}_i = \{f_i(y) : \Re^{a_i+b_i} \mapsto \Re\},$$

for $y = (y_1, \ldots, y_{(a_i+b_i)})$, has the *isolation property* if it contains the function $\tau_j(y) = y_j$ for all $j = 1, 2, \ldots, (a_i + b_i)$. If each $\mathcal{F}_i$ satisfies the isolation property, then the following conditions are directly satisfied.

$$E(f^*) \leq E(S^*)$$

and

$$\hat{E}(\hat{f}) \leq \hat{E}(\hat{S}).$$

The first condition is useful only if f^* can be computed, which in turn requires the knowledge of the distributions. If the distributions are not known, then $\hat{f}$ can be used as an approximation. We now show that with probability $1 - \delta$, we have

$$E(\hat{f}) - E(f^*) \leq \epsilon$$

given a sufficiently large sample, and that the physical law and the fusers classes are Lipschitz.

Theorem 3.1 *Consider that the physical law is Lipschitz with constant k_L, and parameters, estimators and measurements are bounded such that $p \in [-C, C]^n$, $m_i \in [-A, A]^{a_i}$, and $e_i \in [-B, B]^{b_i}$. Let each fuser class $\mathcal{F}_i$ be Lipschitz with constant k_{f_i}. Let*

$$d = \sum_{i=1}^{n} (a_i + b_i)$$

and

$$k = k_L \max(k_{f_1}, k_{f_2}, \ldots, k_{f_n}).$$

Then given a sample size

$$s = \frac{512k(A+B)}{\epsilon^2}\left[d \ln\left(\frac{32k(A+B)}{\epsilon}\right) + \ln(8/\delta)\right],$$

we have

$$\mathbf{P}\left[E(\hat{f}) - E(f^*) > \epsilon\right] \leq \delta,$$

irrespective of the sensor distributions. Furthermore, $E(\hat{f}) \to E(f^*)$, *as* $s \to \infty$.

Proof: Consider the function class

$$\mathcal{L} = \{L(f_1, f_2, \ldots, f_n) : f_1 \in \mathcal{F}_1, \ldots f_n \in \mathcal{F}_n\},$$

where $L(f_1, f_2, \ldots, f_n)$ is defined on $[-A, +A]^d$. Since each f_i is Lipschitz with constant k_{f_i}, $L(f_1, f_2, \ldots, f_n)$ is Lipschitz with constant $k = k_L \max(k_{f_1}, \ldots, k_{f_n})$, and furthermore due to boundedness of estimators $L(f_1, f_2, \ldots, f_n)$ is upper-bounded by $k(A+B)$. By the result of Vapnik [18] (page 41), we have

$$\begin{aligned}&\mathbf{P}\left[E(\hat{f}) - E(f^*) > \epsilon\right] \\ &\quad \leq \mathbf{P}\left[\sup_{l \in \mathcal{L}} |\hat{E}(l) - E(l)| > \epsilon/2\right].\end{aligned}$$

To see this result, consider the condition

$$\mathbf{P}\left[\sup_{l \in \mathcal{L}} |\hat{E}(l) - E(l)| > \epsilon/2\right] < \delta$$

or equivalently

$$\mathbf{P}\left[\sup_{l \in \mathcal{L}} |\hat{E}(l) - E(l)| < \epsilon/2\right] > 1 - \delta.$$

Then, with probability $1 - \delta$, we have

$$\begin{aligned}E(f^*) &\leq \hat{E}(f^*) + \epsilon/2 \\ &\leq \hat{E}(\hat{f}) + \epsilon/2 \\ &\leq E(\hat{f}) + \epsilon\end{aligned}$$

where the first and third inequalities are due to the application of supremum bound for f^* and $\hat{f}$, respectively, and the second inequality is due to the condition $\hat{E}(f^*) \geq \hat{E}(\hat{f})$. As a result, we have

$$\mathbf{P}\left[E(\hat{f}) - E(f^*) > \epsilon\right] \leq \delta,$$

which shows the above Vapnik's result.

Based on Lemma 3 of Appendix of Krzyzak *et al.* [3] (which itself is based on Pollard [7]), we have

$$\begin{aligned}&\mathbf{P}\left[\sup_{l \in \mathcal{L}} |\hat{E}(l) - E(l)| > \epsilon\right] \\ &\quad \leq 8 N_\infty(\epsilon/8, \mathcal{L}) e^{-\frac{\epsilon^2 s}{128k(A+B)}}.\end{aligned}$$

By combining the above two results, we have

$$\begin{aligned}&\mathbf{P}\left[E(\hat{f}) - E(f^*) > \epsilon\right] \\ &\quad \leq 8 N_\infty(\epsilon/16, \mathcal{L}) e^{-\frac{\epsilon^2 s}{512k(A+B)}}.\end{aligned}$$

Thus, the sample size required to ensure the result of the theorem is given by

$$\frac{512k(A+B)}{\epsilon^2}\left[\ln N_\infty(\epsilon/16, \mathcal{L}) + \ln(8/\delta)\right].$$

Now we have

$$N_\infty(\epsilon/16, \mathcal{L}) \leq \left[\frac{32k(A+B)}{\epsilon}\right]^d$$

which proves the first part of the theorem.

To show the asymptotic convergence, let

$$\delta(s, \epsilon) = 8 N_\infty(\epsilon/16, \mathcal{L}) e^{-\frac{\epsilon^2 s}{512k(A+B)}}$$

show the explicit dependence of δ on s and ϵ. Under the finiteness of $N_\infty(\epsilon, \mathcal{L})$, the consistency result follows from the Borel-Cantelli Lemma [1] if

$$\sum_{l=1}^{\infty} 8 N_\infty(\epsilon/16, \mathcal{L}) e^{-\frac{s\epsilon^2}{512k(A+B)}} < \infty$$

for every $\epsilon > 0$. This condition is true since

$$\begin{aligned}\sum_{s=1}^{\infty} e^{-\frac{s\epsilon^2}{512k(A+B)}} &\leq \int_{x=0}^{\infty} e^{-\frac{x\epsilon^2}{512k(A+B)}} dx \\ &\leq \frac{512k(A+B)}{\epsilon^2} e^{-\frac{\epsilon^2}{512k(A+B)}}\end{aligned}$$

which is finite for all $\epsilon > 0$. □

This theorem provides a distribution-free finite sample result: given a sufficiently large sample size, with a probability $1 - \delta$, the cost of the sample-based solution is within ϵ of the lowest achievable cost (which can only be computed if all error distributions are known). This result is distribution-free, i. e. it depends only on the measurements and does not depend on the sensor distributions. It is the best possible result in that stronger results such as showing $\delta = 0$ is not possible, since f^* depends on a distribution (which is not finite-dimensional) and $\hat{f}$ depends on a *finite* sample. Results similar to the asymptotic result shown in the above theorem are more common in the statistics literature [8]. The finite sample result, however, is stronger in that it implies the asymptotic result, and also establishes that the method is justified even for small sample sizes. The actual sample size estimate provided in the above theorem, we believe, can be sharpened.

The smoothness conditions required in this theorem are quite reasonable. The Lipschitz condition is satisfied for a number of physical laws, although not always guaranteed. Similar condition was used in converting the decision fusion rules designed for known distributions to sample-based ones in [10]. The isolation and Lipschitz properties required of the fusers are satisfied in a number of cases such as linear combinations with bounded coefficients and piecewise linear feedforward networks [11].

The following corollary is a weaker version of Theorem 3.1 since $E(f^*) \leq E(S^*) \leq E(\hat{S})$.

Corollary 3.1 *Let* $\mathcal{F}_i$ *satisfy the isolation property for all* $i = 1, 2, \ldots, n$. *Under the same conditions as Theorem 3.1, we have following conditions satisfied.*

$$\mathbf{P}\left[E(\hat{f}) - E(S^*) > \epsilon\right] \leq \delta$$

$$\mathbf{P}\left[E(\hat{f}) - E(\hat{S}) > \epsilon\right] \leq \delta.$$

Informally speaking, this corrollary shows that the error of the computed fuser $\hat{f}$ is not likely to be much higher than that of the best basic set, and could be much smaller. Indeed the above theorem states that $\hat{f}$ will be closer to f^* which can have much smaller error than S^*.

The isolation property was first proposed in [15, 9] for the concept and sensor fusion problems. If $\mathcal{F}_i$ is the set of linear combinations, i. e. $f_i(y) = w_1 y_1 + \ldots + w_k y_k$, for $w_i \in [-W, W]$, this property is trivially satisfied. Lipschitz constant of f_i is W.

Example 2: Consider the scenario in Example 1, and

$$f_2(m_{2,1}, m_{2,2}) = w_1 m_{2,1} + w_2 m_{2,2}.$$

For the choice $w_1 = 0.3$ and $w_2 = 1$, we have $f_2(m_{2,1}, m_{2,2}) = a + 0.3\delta$, and

$$L[m_{1,1}, f_2(m_{2,1}, m_{2,2})] = (\epsilon - 0.3m\delta)^2,$$

which is always smaller than $L[m_{1,1}, m_{2,1}]$, and is smaller than $L[m_{1,1}, m_{2,2}]$ for small $\delta \leq a$ which is true with probability one since δ is a zero-mean random variable. $\square$

4 Methane Hydrates Well Logs

In this section, we briefly describe the methane hydrates problem [4] and show how the proposed fusion method improves the porosity estimation. Gas hydrates are crystalline substances composed of water and gas, in which gas molecules are contained in cage-like lattices formed by solid water. Gas hydrates are present in permafrost regions and beneath the sea in sediment of outer continental margins. Among the several gas hydrates, methane hydrates are believed to be the most common in nature [4]. Methane hydrates represent an alternative to conventional gas as a fuel, and they are expected to be available in volumes exceeding known estimates for conventional gas reserves. One of the challenging problems is to predict the prescence of hydrates using measurements collected at wells located in certain locations such as off the US coast in mid-Atlantic and Mackenzie Delta in Northwest Canada.

At each well, a number of measurements are collected using a suite of sensors. These measurements include density, neutron porosity, acoustic transit-time, and electric resistivity, collected at various depths in the well [2]. Our focus is on the estimation of the *porosity* at various depths. Our data consists of 3045 sets of measurements each collected at different depths in a single well. There are a variety of methods to estimate porosity based on different principles and utilizing different measurements [5]. We employed six known methods for estimating the porosity based on neutron measurements ($\hat{\phi}_1$), density measurements ($\hat{\phi}_2$), fluid velocity equation ($\hat{\phi}_3$), acoustic travel time based on S-wave ($\hat{\phi}_4$), time-average equation based on P-wave ($\hat{\phi}_5$), and Wood's equation ($\hat{\phi}_6$) (see [5, 6] for details on these models).

One of the well-established physical laws relates the parameters of porosity (ϕ), density (ρ), and hydrate concentration (ψ), as follows [5, 6]

$$L[\phi, \psi, \rho] = (\phi[\rho_m - (1-\psi)\rho_w + \psi\rho_h] - \rho + \rho_m)^2 = 0,$$

where ρ_m, ρ_w, and ρ_h are known constants. In this equation, we use the only one measurement for density $\hat{\rho}$ and and a single estimator $\hat{\psi}$ for the hydrate concentration using the Archie's equation [6]. As a first step solution to this problem, we assume that the estimators $\hat{\rho}$ and $\hat{\psi}$ are reasonably accurate, and the objective is to fuse the porosity estimates. When an estimator $\hat{\phi}$ is used in the above equation we obtain a non-zero value in general, so that

$$L[\hat{\phi}, \psi, \rho] = \left(\hat{\phi}[\rho_m - (1-\psi)\rho_w + \psi\rho_h] - \rho + \rho_m\right)^2,$$

is non-zero. Thus, the error due to a set of estimators $\hat{\phi}$, $\hat{\psi}$, and $\hat{\rho}$ is given by

$$\begin{aligned} \hat{E}(\hat{\phi}, \hat{\psi}, \hat{\rho}) &= \sum_{z=1}^{3045} L[\hat{\phi}, \hat{\psi}, \hat{\rho}] \\ &= \Big(\hat{\phi}(z)[\rho_m - (1-\hat{\psi}(z))\rho_w + \hat{\psi}(z)\rho_h] \\ &\quad -\hat{\rho}(z) + \rho_m\Big)^2. \end{aligned}$$

The following are the error estimates for the individual estimators.

estimator	$\hat{E}(.)$
$\hat{\phi}_1$	1.454920
$\hat{\phi}_2$	1.407088
$\hat{\phi}_3$	1.400968
$\hat{\phi}_4$	1.400967
$\hat{\phi}_5$	1.728479
$\hat{\phi}_6$	1.410965
fused estimate	0.079374

Note that in terms of $\hat{E}(.)$, the best estimator for porosity is $\hat{\phi}_4$. Then, we consider a fuser based on the linear combination of the estimators

$$\hat{\phi}_F = w_7 + \sum_{i=1}^{6} w_i \hat{\phi}_i,$$

where $(w_1, \ldots, w_7) \in \Re^7$ is the weight vector. Recall that the isolation property is satisfied by linear combinations. We computed the weight vector to minimize the error

$$\hat{E}(\hat{\phi}_F) = \sum_{z=1}^{3045} L\left[w_7 + \sum_{i=1}^{6} w_i \hat{\phi}_i(z), \hat{\psi}(z), \hat{\rho}(z)\right].$$

The solution is given as follows for the well log data:

w_1	1.198060
w_2	2.871261
w_3	-0.264511
w_4	-7.243817
w_5	2.254488
w_6	-6.330654
w_7	2.458406

The computation is performed using the netlib library routines for solving the least square problems. The error achieved by $\hat{\phi}_F$ is about 20 times better than that of the best estimator $\hat{\phi}_4$. More dramatic performance of the fuser was observed when smaller subsets of well log data were utilized. However, since the magnitude of the error depends critically on the exact form of $L[.]$, the improvement must be interpreted in the context, as described at the end of Section 2. Nevertheless, the fused estimate is significantly better than any individual estimator for the porosity.

5 Conclusions

We presented an information fusion method that applies to physical systems wherein accurate measurements of physical parameters are not possible. The fusion operation is achieved by utilizing the physical laws that relate various parameters of the system. Compared to the traditional information fusion problems, this formulation is a significant departure in that it lacks the ground truth in the form of training sample. We presented a methodology that utilizes physical laws to compare various sets of measurements and estimators. If the error distributions are known, the best set of measurements and estimators can be computed. In practice, the distributions are not known but only measurements are available. We presented a fusion method that combines various measurements and estimators to achieve performance at least as good as the best set of measurements. Such fuser again depends on the sensor distributions, and cannot be computed with probability one if only measurements are available. We showed that a close approximation to the this optimal fuser can be computed such that with a high probability the solution performs at least as good as the best set of measurements, given large enough sample size. We illustrated the effectiveness of this method for the data fusion for the well log data in methane hydrates exploration.

The proposed fuser is based on the isolation property. The projective fusers proposed recently [13] are more powerful in that they perform at least as good as the best subset of estimators, but are harder to compute based on a sample. Also, metafusers that combine various fusers to ensure that the best performance of the fusers as well as estimators have been proposed [13]. The study of projective fusers and metafusers for the proposed formulation will be of future interest. Furthermore, a more detailed analysis of the methane hydrates data by using a number of estimators for parameters such as density and concentration, and also the other types of fusers, will be of future interest.

Acknowledgements

This research is sponsored by the Engineering Research Program of the Office of Basic Energy Sciences, U.S. Department of Energy, under Contract No. DE-AC05-00OR22725 with UT-Battelle, LLC, and the Office of Naval Research under order N00014-96-F-0415.

References

[1] P. Billingsley. *Probability and Measure.* John Wiley and Sons, New York, second edition, 1986.

[2] S. R. Dallimore, T. Uchida, and T. S. Collett, editors. *Scientific Results from JAPEX/JNOC/GSC Mallik 2L-38 Gas Hydrate Research Well, Mackenzie Delta: Geological Survey of Canada Bulletin 544.* Geological Survey of Canada, Bulletin 544, 1999.

[3] A. Krzyzak, T. Linder, and G. Lugosi. Nonparametric estimation and classification using radial basis function nets and empirical risk minimization. *IEEE Transactions on Neural Networks*, 7(2):475–487, 1996.

[4] K. A. Kvenvolden. A primer on the geological occurrence of gas hydrate. In J. P. Hemriet and J. Mienert, editors, *Gas Hydrates: Relevence to World margin Stability and Climate Change*, volume 137, pages 9–30. Geological Society, London, Special Publication, 1998.

[5] G. Mavko, T. Mukerji, and J. Dvorkin. *The Rock Physics Handbook: Tools For Seismic Analysis in Porus Media.* Cambridge University Press, 1998.

[6] M. Miyairi, K. Akihisa, T. Uchida, T. S. Collett, and S. R. Dallimore. Well-log interepretation of gas-hydrate-bearing formations in the JAPEX/JNOC/GSC Mallik 2L-38 gas hydrate research well. In *Scientific Results from JAPEX/JNOC/GSC Mallik 2L-38 Gas Hydrate Research Well, Mackenzie Delta: Geological Survey of Canada Bulletin 544*, pages 281–293, 1999.

[7] D. Pollard. *Convergence of Stochastic Processes.* Springer-Verlag, New York, 1984.

[8] B. L. S. Prakasa Rao. *Nonparametric Functional Estimation.* Academic Press, New York, 1983.

[9] N. S. V. Rao. Fusion methods for multiple sensor systems with unknown error densities. *Journal of Franklin Institute*, 331B(5):509–530, 1994.

[10] N. S. V. Rao. Distributed decision fusion using empirical estimation. *IEEE Transactions on Aerospace and Electronic Systems*, 33(4):1106–1114, 1996.

[11] N. S. V. Rao. A fusion method that performs better than best sensor. In *First International Conference on Multisource-Multisensor Data Fusion.* 1998. 19-26.

[12] N. S. V. Rao. Multiple sensor fusion under unknown distributions. *Journal of Franklin Institute*, 336(2):285–299, 1999.

[13] N. S. V. Rao. Projective method for generic sensor fusion problem. In *IEEE/SICE/RSJ International Conference on Multisensor Fusion and Integration for Intelligent Systems*, pages 1–6. 1999.

[14] N. S. V. Rao. Simple sample bound for feed-forward sigmoid networks with bounded weights. *Neurocomputing*, 29:115–122, 1999.

[15] N. S. V. Rao, E. M. Oblow, C. W. Glover, and G. E. Liepins. N-learners problem: Fusion of concepts. *IEEE Transactions on Systems, Man and Cybernetics*, 24(2):319–327, 1994.

[16] N. S. V. Rao, D. B. Reister, and J. Barhen. Fusion method for physical systems based on physical laws. In *Proceedings of 3rd International Conference on Information Fusion*, 2000.

[17] L. G. Valiant. A theory of the learnable. *Communications of the ACM*, 27(11):1134–1142, 1984.

[18] V. Vapnik. *Estimation of Dependences Based on Empirical Data.* Springer-Verlag, New York, 1982.

[19] V. N. Vapnik. *The Nature of Statistical Learning Theory.* Springer-Verlag, New York, 1995.